

APPENDIX N

Proc continuous-process(x)

```

// x is a data entry/value
If the corresponding variable is not selected by AMB, return;
If x is a missing value
    Mark this record invalid;
    Substitute it with mean value;
    //mean value of this variable in training set collected during AMB
Else
    If x > max    // maximum value of this variable in training set collected during AMB
        x = max;
        Mark this record invalid;
    End If

    If x < min    // minimum value of this variable in training set collected during AMB
        x = min;
        Mark this record invalid;
    End If
End If

```

If the corresponding variable is exponentially distributed

```

Retrieve the mean and min value for log-scaling;
// It is mean and minimum value of samples of this predictor in training set when conduct
// exponential distribution test, might be different from those in whole training set

```

$$x = 1 - e^{-\frac{x-\min}{\text{mean}-\min}};$$

End If

Retrieve the mean and norm value for normalization;

$$x = \frac{x - \text{mean}}{\text{norm}};$$

Put x in the design matrix according to its column index and row number.

Proc categorical-process(x)

```
// x is a data entry/value, m is the number of records
```

If the corresponding dummy is not retained in the model then Return;

```
    Get the column index of this categorical variable in the design matrix [i:j];
    //1<=i<=j;
```

Fill 0s in entry(ies)[m, i:j];

If this dummy appears in the training set

Get the column index of this dummy, k (i <= k <= j, or k < 0);

If k > 0

Fill a 1 in entry (m,k);

End If

Else

Mark this record invalid;

End If
For k = i:j
 x = value of entry (m, k) ; //1 or 0
 Get the mean and norm value for normalization;
 $x = \frac{x - \text{mean}}{\text{norm}}$;
 entry $(m, k) = x$;
End For